Running head: EEG-BASED ESTIMATION OF MENTAL FATIGUE

EEG-based Estimation of Mental Fatigue

Leonard J. Trejo[*a], Rebekah Kochavi[a], Karla Kubitz[b],

Leslie D. Montgomery[a], Roman Rosipal[a], Bryan Matthews[a]

[a]NASA Ames Research Center, MS 269-3, Moffett Field, CA, USA 94035-1000

[b]Towson University, 8000 York Rd., Towson, Md., 21252

[*] ltrejo@mail.arc.nasa.gov; phone 1-650-604-2187, fax 1-650-604-4036; ic.arc.nasa.gov

Abstract

Mental fatigue from continuous mental arithmetic is associated with an across-subjects effect of increased frontal theta and parietal alpha rhythms. A statistical classifier can use this effect to model EEG-fatigue relationships. Subjects viewed 4-digit problems on a computer, solved the problems, and pressed keys to respond. They performed until either they felt exhausted or 3 hrs had elapsed. Pre- and post-task mood scales confirmed that fatigue increased and energy decreased. We examined accuracy, response times (RT), amplitudes of ERP components N1, P2, and P300, and power of frontal theta and parietal alpha EEG rhythms over time. Mean RT rose from 6.7 s to 7.9 s but accuracy did not change significantly. The effects of time on ERP component amplitudes or latencies were inconclusive. Mean frontal EEG theta and alpha power rose by 29% and 44%, respectively. We also used EEG frequency spectra to model the effects of time using a kernel partial least squares (KPLS) classifier, classifying EEG segments as being from the first or last 15 minutes. Accuracies ranged from 91% to 100% correct.

EEG, ERP, mental fatigue, alertness, situation awareness, performance monitoring

Measures and Models for Predicting Mental Fatigue

There are countless occupations today, including many in aviation, transport, aerospace, military, medicine, and industrial settings, in which fatigued individuals routinely operate complex, automated systems. This undesirable state of affairs has contributed to more than a few well-publicized—and many not so well publicized—disasters. For example, operator fatigue was involved in the Three-Mile Island and Chernobyl nuclear power plant disasters, the EXXON VALDEZ grounding, and the Guantanamo Bay airline crash (Fatigue Resource Directory, 2000; Dinges, 1995). "There is [also] much evidence that [operator] fatigue has contributed to serious incidents and accidents in industrial operations … and to accidents in <u>all</u> [emphasis added] modes of transportation" (Dinges, 1995, p. 8). Given the prevalence, and the potentially serious outcomes, of this problem, there has been, and still is, much scientific interest in operator fatigue. There has also been much effort devoted to evaluating the effects of operator fatigue and to monitoring operator state (Gevins et al., 1995; Kennedy, 1953; Wilson & Fisher, 1991; 1995; Russo, Stetz, & Thomas, 2005).

Although fatigue can occur because of either mental or physical activity (Stedman's 2005), in this study we focused on mental fatigue because it has many serious, harmful effects on cognitive functioning. Mental fatigue decreases both vigilance (Mast & Heimstra, 1964; Šípoš, 1971; Smit, Eling, & Coenen, 2004) and secondary task performance (van der Hulst, Meijman, & Rothengatter, 2001). It also hampers problem solving (Horne, 1988; Van Der Linden, Frese, & Meijman, 2003; Van Der Linden, Frese, & Sonnentag, 2003), lessens situation awareness (Vidulich, Stratton, Crabtree, & Wilson, 1994), and decreases risk and effort tolerance (Barth, Holding & Stamford, 1976; Brown, Tickner, & Simmonds, 1970; Neri, Shappell, & DeJohn, 1992; Shingledecker & Holding, 1974). We defined mental fatigue as the unwillingness of alert, motivated subjects to continue performing mental work (Montgomery, Montgomery, & Guisado, 1995). In this way, mental fatigue differs from the other factors that also influence operator functioning, including sleepiness, lack of motivation, monotony, learning, and physical fatigue.

We approached the problem of mental fatigue from two complementary approaches, human psychophysiology and neuroengineering. First, we show the effects of mental fatigue, elicited by sustained mental work in a laboratory setting, on both endogenous electroencephalograms (EEG) and on event-related brain potentials (ERP). This is noteworthy because the simultaneous measurement of EEGs and ERPs tells us about the effects of mental fatigue on ongoing brain processes and about the brain's responses to particular events. We focused our data collection and analysis efforts on frontal midline theta band activity, on parietal alpha band activity, and on the N100, P200, and P300 ERPs because they respond systematically to changes in operator state (Gevins et al., 1998; Kramer, Trejo, & Humphrey, 1995; Trejo, Kramer, & Arnold, 1995). Second, we show the feasibility of modeling and predicting the effects of mental fatigue at sub-minute time scales in individual subjects. That is, we developed and cross-validated statistical learning algorithms for classifying segments of only a few seconds worth of EEG activity according to fatigue. We think this is important because it represents a thorough and novel use of robust statistical methods to model the relationships between mental fatigue and EEGs across time in individual subjects.

There are several important differences between our work and previous studies. First, most of the previous studies have examined the effects of either monotony (using vigilance-like paradigms) or of mental workload (using mostly short duration, high workload experimental sessions to examine fatigue). In contrast, we used a sustained low-workload, mental arithmetic task and encouraged subjects to preserve alertness, motivation, and high response accuracy to lessen monotony or task difficulty-related effects. Second, nearly all the previous studies (especially the few that examined mental fatigue), focused either on changes in ongoing EEG activity or on changes in ERPs and not on both variables simultaneously. We examined both aspects of cognitive functioning simultaneously and continuously throughout sustained task performance. Also, we examined the effects of mental fatigue on the parts of the EEG and ERPs that previous studies have shown to be consistently responsive to changes in operator state. Third, with some exceptions (Duta, Alford, Wilson, & Tarassenko, 2004; Gevins & Smith, 1999; Gevins et al., 1998; Nikolaev, Ivanitskii, & Ivanitskii, 1998; Trejo & Shensa, 1999; Venturini, Lytton, & Sjnowski, 1992; Wilson & Russell, 2003a; Wilson & Russell, 2003b), most of the previous studies used signal processing and analytical techniques of somewhat limited power and flexibility, for example, analyzing EEG spectra with low frequency resolution and classifying operator state using either multiple regression or linear discriminant analyses. Instead, we analyzed high-resolution EEG power spectra in overlapping intervals synchronized with task stimuli. In addition, to improve classification accuracy, we used a classifier developed from statistical learning theory to model and predict mental fatigue.

This report will include an assessment of the studies that have used EEG and ERP activity as a means to evaluate the effects of mental fatigue or other conditions on operator state. Most of these studies either focused on evaluating the effects of monotony or mental workload on EEG and ERP activity. A few focused on evaluating the effects of mental fatigue, in a fashion similar to ours, on EEG and ERP activity. This report will also include a discussion of the studies that have used one or more EEG correlates of mental fatigue to try to classify operator state. Finally, we will detail our methods and findings and discuss them in the context of previous research.

Researchers interested in using EEG and ERP activity as a means to evaluate the effects of mental fatigue or other conditions on operator state have typically examined the effects of specific experimental manipulations on EEG activity (and related variables). Although there is certainly conceptual overlap (i.e., because it can be argued that doing almost anything long enough can lead to mental fatigue), most of these studies have manipulated the constructs of vigilance, mental workload, or mental fatigue. The subjects used in these studies have varied widely, ranging from college student volunteers to pilots, air traffic controllers, and radar operators. The tasks used in these studies have been similarly diverse, ranging from fairly short to quite long duration tasks (i.e., from 15 min to over 15 hours) and from controlled laboratory tasks to real or simulated operations. Examples of the former include the working memory task studied by Gevins and colleagues (e.g., the n-back task, Gevins & Smith, 2000; Gevins et al., 1998; Gevins, Smith, McEvoy & Yu, 1997) and the 'switching' task studied by Lorist and his colleagues (Lorist, Boksem, & Ridderinkhof, 2005; Lorist et al., 2000). Examples of the latter include the sonar monitoring tasks studied by Makeig and his colleagues (Jung, Makeig, Stensmo, & Sejnowski, 1997; Makeig & Inlow, 1993; Makeig & Jung, 1995) and the flight/ flight simulator tasks favored by Wilson and his colleagues (Hankins & Wilson, 1998; Wilson & Russell, 2003b). The dependent variables used in these studies have included recordings of spontaneous EEG activity (i.e., in an assortment of frequency bands and sites) and of various ERPs (including the N100, P200, P300, and contingent negative variation or CNV).

The most consistent finding in these studies is that fatigue-related manipulations are associated with increased theta power at the midline frontal location (i.e., Fz) and decreased alpha band power at one or more parietal locations (e.g., P7 and P8) (Fairclough, Venables, & Tattersall, 2004; Gevins & Smith, 1999; Gevins et al., 1997; Gundel & Wilson, 1992; Hankins & Wilson, 1998; Laukka, Järvilehto, Alexandrov, & Lindqvist , 1995; Makeig & Inlow, 1993; Mecklinger, Kramer, & Strayer, 1992; Smith, Gevins, Brown, Karnik, & Du, 2001; Smith, McEvoy, & Gevins, 2002; Pellouchoud, Smith, McEvoy, & Gevins, 1999; Vidulich et al., 1994). It has also been consistently shown that ERP components such as N100 and P300 that are associated with primary task stimuli increase in amplitude along with increases in attentional demand, task difficulty, or mental workload (Mangun & Hillyard, 1990; Sirevaag, Kramer, Coles & Donchin, 1989; Trejo et al., 1995; Ullsperger, Metz, & Gille, 1988; Wickens, Kramer, Vanasse, & Donchin, 1983) and that ERP components not associated with primary task stimuli (i.e., either secondary task stimuli or task irrelevant stimuli) do the opposite (Haga, Shinoda, & Kokinbun, 2002; Humphrey & Kramer, 1994; Israel, Chesney, Wickens, & Donchin, 1980; Isreal, Wickens, Chesney, & Donchin, 1980; Kramer, Sirevaag, & Hughes, 1988; Kramer et al., 1995; Kramer, Wickens, & Donchin, 1983; Wilson, Fullenkamp, & Davis, 1994). Although there have been relatively few studies focusing only on mental fatigue, ERP amplitudes appear to decrease with mental fatigue (Lorist et al., 2005; Lorist et al., 2000; Smith et al., 2002).

The data processing and the analytical approaches used in these studies have changed in several ways over the past two decades. First, early studies tended to examine the effects of fatiguing manipulations on the traditional EEG frequency bands (e.g., the delta, alpha, theta, and beta bands). More recent studies have demonstrated the effectiveness of examining one or more specific, EEG frequency bands at particular scalp locations (e.g., 5-7 Hz activity at site Fz; Gevins et al., 1998; Gevins et al., 1997; Laukka et al., 1995; Smith et al., 2001). In addition, several researchers have found it advantageous to use principal component analysis to derive frequency bands (Makeig & Jung, 1995; Wilson & Fisher, 1995). Second, the early studies (and some of the more recent studies) usually analyzed the EEG in the frequency domain, using the Fast Fourier transform (e.g., Belyavin & Wright, 1987; Gevins, Zeitlin, Ancoli, & Yeager, 1977; Makeig & Inlow, 1993). More recent studies have found that more information is obtained by analyzing the EEG using time series analyses (e.g., autoregressive or moving average models, Duta et al., 2004; Florian & Pfurtscheller, 1995) or time-frequency analyses (e.g., wavelet models, Trejo & Mullane, 1995; Trejo & Shensa, 1999). Third, early studies used multiple regression and, occasionally, linear discriminant-based analyses to attempt to predict operator functioning (e.g., Makeig, Jung, & Sejnowski, 1995; Kramer, Trejo & Humphrey, 1996; Trejo & Mullane, 1995; Trejo & Shensa, 1999; Wilson & Fisher, 1991; Wilson & Russell, 2003a). More recent studies have demonstrated the added value of using statistical learning methods such as neural networks to classify operator state (Duta et al., 2004; Gevins & Smith, 1999; Gevins et al., 1998; Nikolaev et al., 1998; Trejo & Shensa, 1999; Venturini et al.,1992; Wilson & Russell, 2003a; Wilson & Russell, 2003b). In fact, two studies (Trejo & Shensa, 1999; Wilson & Russell, 2003a) compared linear and neural-network classifiers and found that the neural network-based classifiers had the highest classification accuracies and the best generalization.

Researchers interested in using one or more EEG correlates of mental fatigue to try to classify operator state have typically administered a stimulus, recorded EEG activity, and then tried to classify EEG activity into one or more ‘states’ (fatigued/ not fatigued/ high mental workload/ low mental workload). Gevins et al. (1977) were among the first to attempt to develop an online, EEG-based drowsiness detector. Based on automated sleep scoring research, they developed a computer program to predict operator state (i.e., as fatigued or as not fatigued).

The program did a spectral analysis on the EEG data and calculated the ratios of delta to alpha, and theta to alpha activity. Calculated ratios were compared to 'drowsiness threshold' ratios previously calculated and these comparisons were used to predict operator state. Gevins et al. tested their computerized drowsiness detector on EEG recordings from 31 individuals and found that 92% of the training, and 84% of testing, epochs were identified as drowsy both by expert scorers and by the drowsiness detector. Thus, the computer program was able to predict drowsiness based on recordings of EEG activity with a fairly high degree of accuracy. Questions remained about how to optimize the accuracy and immediacy (i.e., make more 'real time') of these predictions.

Many of the early efforts at classification involved the use of stepwise linear discriminant analysis (Kramer et al., 1996; Makeig et al., 1995; Trejo & Mullane, 1995; Trejo & Shensa, 1999; Wilson & Fisher, 1991; Wilson & Fisher, 1995; Wilson & Russell, 2003a) to predict performance from the EEGs. More recent efforts have either involved the development of some sort of individualized task engagement index (Smith et al., 2001; Trejo et al., 1995) or the use of neural network-based algorithms (Duta et al., 2004; Gevins et al., 1998; Gevins & Smith, 1999; Jung et al., 1997; Makeig et al., 1995; Nikolaev et al., 1998; Smith et al., 2002; Trejo & Shensa, 1999; Venturini et al., 1992; Wilson & Russell, 2003a; Wilson & Russell, 2003b). This change is driven by the need to be able to monitor operator state over very short periods of time (hence small chunks of data) and to cope with nonstationarity or complexity in the parameters of the distributions of EEG spatial and spectral measures.

The most common findings from the classification studies are that both EEG and ERP activity can be used to monitor mental workload with fairly high levels of accuracy. For example, Gevins et al. (1998) used neural networks to classify EEG activity from two different workload levels (i.e., high and low) with 98% accuracy. Nikolaev et al. (1998) used neural networks to classify EEG activity from two different types of tasks with 89% accuracy (for the complex tasks). They also found that alpha activity at the occipital and parietal sites was the most heavily weighted. EEG appears to have the greatest potential for 'real time' monitoring because ERPs still need to be averaged over a number of occurrences. However, studies have shown that smaller amounts of EEG activity and shorter, fewer ERPs included in the average can be used as classifiers. Moreover, there have also been efforts to use bootstrapping to determine how much ERP data is needed to be able to discriminate operator state (Humphrey & Kramer, 1994).

There are several important implications for future studies to be drawn from this work. First, classification algorithms must be both individualized and multivariate (Galbraith & Wong, 1993; Smith et al., 2001). Second, machine-learning algorithms such as neural networks often classify more reliably than simple regression or linear discriminant function-based classification algorithms. Third, smaller chunks of EEG data and ERP data can be used to predict operator state. Although, prediction accuracy is related to the amount of ERP data, using more sites increased prediction accuracy even for smaller segments of EEG/ ERP data (Kramer et al., 1996). Fourth, some type of saliency analysis is important in order to determine which aspects of the EEG yield the most accurate predictions (Wilson & Russell, 2003a; 2003b). Wilson and Russell (2003a; 2003b) found that classification accuracy went from to 88% to 90% after saliency analysis. Finally, the bands and sites that contribute the most to the prediction algorithms vary depending on the task and on the subject.

In summary, we tested several hypotheses in this study. First, we tested the hypothesis that the relative power of theta, alpha, and other EEG rhythms would co-vary with the degree of fatigue that subjects experienced. Specifically, we tested the hypotheses that theta activity would increase and that alpha activity would decrease across time on task. In addition, we tested the null hypothesis that EEG power in specific theta and alpha bands would remain constant over the course of a fatigue-inducing task. Second, we tested the hypothesis that the ERPs would be responsive to mental fatigue. Specifically, we hypothesized that the visual N100 component would decrease in amplitude and increase in latency with fatigue. Our hypotheses about the P300 were more theory based. We hypothesized one of two possibilities. That is, that the P300 would either increase in amplitude and latency or that the P300 would decrease in amplitude and latency. The former suggests mental effort-like effects and the latter suggests monotony or sleep deprivation-like effects. In addition we tested the null hypotheses that specific ERP components, that is, the N100, P200, and P300 would remain constant in amplitude and latency. Third, we tested the hypothesis that we could accurately model and predict fatigue as a function of EEG measures using a statistical learning theory based classifier.

Methods

Participants

Data were collected from 33 individuals recruited from the NASA Ames Research Center community. However, 17 of the 33 participants were excluded from analyses. Eight were excluded because their EEG data contained high noise levels, which could not be filtered or corrected. Four were excluded because they either fell asleep (n = 3) or violated experimental protocol (n = 1; wore a watch). Five were excluded because their response times were extremely slow (and consequently they provided too few EEG epochs for analysis). The remaining 16 participants included 12 males and 4 females with a mean age of 26.9 (SD = 7.4) years. All participants signed an informed consent approved by the NASA Ames Research Center and were paid for their participation. Also, according to their self-reports, all of the participants had normal vision and hearing and 14 of the 16 participants were right-handed.

Experimental Design

We tested several hypotheses about the association of subjective moods, observed behavior, performance, and physiological measures during continuous performance of mental arithmetic for up to three hours. We manipulated a single factor, that is, time on task, and used a repeated measures design. Subjective moods were indexed by the Activation Deactivation Adjective Checklist (AD-ACL; Thayer, 1986: 1989) and the Visual Analogue Mood Scales (VAMS; Stern, 1997) questionnaires. Observed behavior included ratings of activity and alertness from videotaped recordings of each participant's performance. The performance measures were response time and response accuracy. The physiological measures were derived from spontaneous EEG and ERPs, including: a) theta activity at Fz (both average power and peak amplitude in the theta band), b) alpha activity at Pz (both average power and peak amplitude in the alpha band), c) peak amplitudes and latencies of the N100, P200, and P300 components of ERPs elicited by onset of the task stimuli

Mental Arithmetic Task

Participants sat in front of a computer with their right hand resting on a 4-button keypad (Neuroscan STIM pad, Compumedics USA, El Paso, TX). Arithmetic summation problems, consisting of four randomly generated single digits, three operators, and a target sum (e.g., 4 + 7 – 5 + 2 = 8), were displayed on a computer monitor (cathode ray tube) continuously until the subject responded (Fig. 1). Only addition and subtraction were used, and equations for which answers were obvious (such as those including several repeated digits) were excluded. The participants: a) solved the problems, b) decided whether their 'calculated sums' were less than, equal to, or greater than the target sums provided, c) indicated their decisions by pressing the appropriate key on the keypad. The keypad buttons were labeled "<," "=," and ">," respectively. Subjects were instructed to answer as quickly as possible without sacrificing accuracy. After a response, there was a 1 s inter-trial interval, during which the monitor was blank. Participants performed the task until either they quit from exhaustion or three hours had elapsed.

Activation Deactivation Adjective Checklist

Thayer's AD-ACL (Thayer, 1986; 1989) is a multi-dimensional checklist reflecting perceptions of activation. Individuals respond to 20 items using a 4-point rating scale (definitely feel, feel slightly, cannot decide, and definitely do not feel). The scoring procedure includes four subscales: energy (reflects general activation), tiredness (reflects general deactivation), tension (reflects high preparatory arousal), and calmness (reflects low preparatory arousal). The AD-ACL is a reliable and valid subjective method (Thayer, 1986; 1989).

Visual Analogue Mood Scales

The VAMS (Stern, 1997) measure eight specific mood states, including afraid, confused, sad, angry, energetic, tired, happy, and tense. The VAMS have a neutral schematic. That is, they have a "mood-neutral" face (and word) at the top of a 100 mm vertical line and they have a "mood-specific" face (and word) at the bottom of the line. Individuals mark the point along the line that best illustrates how they feel at present. Scores range from 0 to 100, with 100 indicating the maximum level of the mood and 0 indicating the minimum level of a mood. Like the AD-ACL, the VAMS are also reliable and valid (Nyenhuis, Stern, Yamamoto, Luchetta, & Arruda, 1997; Stern, 1997).

Observed Activity and Alertness

Activity and alertness were measured by visual inspection of videotapes of each participant's performance. The videotapes showed combined overall scene- and facial views of the participants. For each 15 min interval, a single rater judged levels of alertness and activity (unnecessary motion) on a five-point scale (Table 1). The alertness scale considered the subjective appearance of sleep, sleepiness, dozing off, distraction, yawning, and general alertness. The activity scale considered the frequency of subject motion, including moving in the chair, fidgeting, tapping, or shaking. These ratings were tested for correlations with response time, accuracy, and EEG spectral measures.

EEG Activity

EEG activity was recorded continuously using 32 Ag/AgCl electrodes embedded in an elastic fabric cap (i.e., a Quik-Cap™, Compumedics USA, El Paso, TX). The electrode cap was placed on the participant according to the manufacturer's instructions. The reference electrodes were averaged mastoids and the ground electrode was located at AFz. Vertical and horizontal electrooculograms (VEOG and HEOG) were recorded using bipolar pairs of 10 mm Ag/AgCl electrodes (i.e., one pair superior and inferior to the left eye; another pair to the right and to the left of the orbital fossi). Impedances were maintained at less than 5k Ω for EEG electrodes and 10 kΩ for EOG electrodes. The EEG was amplified and digitized with a 64-channel Neuroscan Synamps™ system (Compumedics USA, El Paso, TX), with a gain of 1,000, sampling rate of 500 s^{-1} and a pass band of 0.1 to 100 Hz. Amplifiers were calibrated with a 50 μV signal prior to each testing session. The signals were digitized and stored on hard disk drives by a computer equipped with Neuroscan Scan 4.2 software (Compumedics USA, El Paso, TX) and archived on optical media (CD-R).

Procedures

Participants: a) were given an orientation to the study, b) read and signed an informed consent document, c) completed a brief demographic questionnaire (age, handedness, hours of sleep, etc.), d) practiced the mental arithmetic task for 10 minutes, and e) were prepared for data collection by having the electrode cap, EOG, and reference electrodes applied. They then completed the pretest self-report measures (i.e., the AD-ACL and VAMS) and performed the mental arithmetic task until either three hours had elapsed or volitional exhaustion had occurred. Task termination was followed by the completion of post-test self-report measures and participant debriefing.

Data Processing

The EEGs, initially processed using Neuroscan Scan 4.2 Edit™ (Compumedics USA, El Paso, TX) software, were: a) submitted to an algorithm for the detection and elimination of eye-movement artifact, b) visually examined and blocks of data containing artifact were manually rejected, c) epoched around the stimulus (i.e., from –5 s pre-stimulus to +8 s post -stimulus), d) low pass filtered (50 Hz; zero phase shift; 12 dB/octave roll off), and e) submitted to an automated artifact rejection procedure (i.e., absolute voltages > 100 μV). The overall single-epoch rejection rate was 47%. The 'cleaned and filtered' epochs were decimated to a sampling rate of 128 Hz. EOG artifact was removed by using wavelet-denoised VEOG and HEOG signals as predictors of the artifact voltages at each EEG electrode in a multivariate linear regression. The residuals of these predictions served to estimate the artifact-free EEG. EEG power spectra were estimated with Welch's periodogram method at 833 frequencies from 0-64 Hz. Peak and average power in the theta and alpha bands were measured at electrodes Fz and Pz, respectively.

The ERP data were initially processed using the same methods as for the EEGs, then epoched around the stimulus (–1.5s pre- to +2 s post-stimulus). The overall rejection rate was

19% for the ERP data. The 'cleaned and filtered' epochs were: a) decimated to a sampling rate of 128 Hz, b) corrected for 'residual' EOG artifact, and c) averaged across the 1st 100, middle 100, and last 100 artifact-free trials ordered according to time on task. The average ERPs were used to measure the latencies and amplitudes of the N100, P200, and P300 components. Latencies were peak latencies and were determined based on visual examination of the spatial distribution for the component (i.e., N100, P200, and P300). Amplitudes were calculated as mean amplitudes (at O2, Fz, and CPz for the N100, P200, and P300 components, respectively) in a window extending +/- 50 ms around the peak latency. Amplitudes were normalized using the McCarthy and Wood (1985) procedure.

Classification Procedures

We classified single EEG epochs using kernel partial least squares, or KPLS-DLR, decomposition of multichannel EEG spectra coupled with a discrete-output linear regression classifier (Rosipal, Trejo, & Matthews, 2003). Through extensive side-by-side testing of EEG data, we found that KPLS-DLR is just as accurate as KPLS-SVC, which uses a support vector classifier for the classification step. KPLS selects the reduced set of orthogonal basis vectors or "components" in the space of the independent variables (EEG spectra) that maximizes covariance with the experimental conditions. DLR finds the linear hyperplane in the space of KPLS components that separates the classes. In a pilot study, and in our present data, we found that the first 15 minutes on task did not produce mental fatigue, whereas mental fatigue was substantial in the final 15 minutes. So we randomly split EEG epochs from the first and last 15-min periods into equal-sized training and testing partitions for classifier estimation. Only the training partition was used to build the final models. The number of KPLS components in the final models was set by five-fold cross-validation. The criterion for KPLS model selection was the minimum classification error rate summed over all (five) cross-validation subsets.

Statistical Analyses

The data were analyzed using either singly or, when appropriate, doubly multivariate repeated measures analyses of variance with either time of measurement (for the self-report, behavior, and EEG analyses) or number of artifact-free trials as a within-subjects factor (for the ERP analyses). The AD-ACL subscale scores (energy, tension, calmness, and tiredness), VAMS subscale scores (afraid, confused, sad, angry, energetic, tired, happy, and tense), behavioral observation data (observed activity and alertness), theta activity data (peak and band-average amplitudes), alpha activity data (peak and band-average amplitudes), N100 data (amplitudes and latencies), P200 data (amplitudes and latencies), and P300 data (amplitudes and latencies) were analyzed using doubly multivariate analyses. Response time and accuracy were analyzed using singly multivariate analyses of variance. For the doubly multivariate analyses, significant multivariate F-ratios were decomposed using single degree of freedom within-subjects contrasts. For the singly multivariate analyses, Huynh-Feldt-corrected degrees of freedom and p-values were reported (i.e., because of sphericity). In both cases, partial η^2 values were reported as effect size estimators.

RESULTS

Self-report Analyses

The AD-ACL subscale scores were analyzed in a doubly multivariate ANOVA with time of measurement (i.e., pretest vs. posttest) as a within-subjects factor. The main effect of time of measurement was significant ($F(4,5) = 10.4$, $p<.01$, h2 = .89). Within-subjects contrasts showed significant linear trends for energy ($F(1,8) = 6.46$, $p<.04$, $\eta^2 = .45$), calmness ($F(1,8) =$

21.3, $p<.002$, $\eta^2 = .73$), and tiredness ($F(1,8) = 6.38$, $p<.04$, $\eta^2 = .44$). Energy decreased from a pretest mean of 12.0 (SD = 4.1) to a posttest mean of 8.6 (SD = 3.7). Calmness decreased from a pretest mean of 16.8 (SD = 1.6) to a posttest mean of 14.1 (SD = 1.8). Tiredness increased from a pretest mean of 10.1 (SD = 4.3) to a posttest mean of 15.3 (SD = 5.7). There was also a non-significant linear trend for tension ($F(1,8) = .92$, $p = .37$). Thus, the AD-ACL data indicate that our manipulation decreased general activation (i.e., self-reported energy) and preparatory arousal (i.e., self-reported calmness) and increased general deactivation (i.e., self-reported tiredness).

The VAMS subscale scores (i.e., for afraid, confused, sad, angry, energetic, tired, happy, and tense) were analyzed in a doubly multivariate ANOVA with time of measurement (i.e., pretest vs. posttest) as a within-subjects factor. The main effect of time of measurement was non-significant (multivariate $F(8,1) = 1.31$, $p = .59$). This analysis suggests that our manipulation, despite its effects on activation and arousal, did not influence moods.

Behavior Analyses

The behavioral observations (i.e., observed activity and alertness) were analyzed in a doubly multivariate ANOVA with time of measurement (i.e., 10 15 min periods) as a within-subjects factor. The main effect of time of measurement was significant ($F(18,178) = 3.70$, $p<.0005$, $\eta^2 = .27$). This analysis suggests that time on task influenced behavior (i.e., observed activity and alertness levels). Moreover, time on task had a progressive effect on behavior. Within-subjects contrasts showed a linear decrease in alertness ($F(1,10) = 10.4$, $p<.009$, $\eta^2 = .51$) and a linear increase in activity ($F(1,10) = 5.88$, $p<.04$, $\eta^2 = .51$). Alertness decreased from a mean of 5.00 (SD = 0.00) in the first 15 min period to a mean of 2.43 (SD = 0.98) in the last 15 min period. Activity increased from a mean of 1.36 (SD = 0.51) to a mean of 2.45 (SD = 1.30), respectively.

The response times (RT) were analyzed in an ANOVA with time of measurement (i.e., 15 min periods) as a within-subjects factor. The main effect of time of measurement was significant (Huynh-Feldt corrected $F(3,39) = 3.78$, $p<.03$, $\eta^2 = .24$). This analysis suggests that time on task influenced performance. Moreover, time on task had a progressive effect on performance (Fig. 2). Within-subjects contrasts showed a significant linear increase in RT ($F(1,12) = 8.29$, $p<.01$, $\eta^2 = .41$) rising from a mean of 6.70 s (SD = 2.18) in the first 15 min period to a mean of 7.87 s (SD = 2.64) in the last 15 min period. We found the same pattern of significant effects for RT analyzed in an ANOVA with fraction of artifact-free trials (i.e., 1st100, middle 100, and last 100) as a within-subjects factor.

Response accuracy was analyzed in an ANOVA with time of measurement (i.e., ten 15-min periods) as a within-subjects factor. The main effect of time of measurement was not significant (Huynh-Feldt corrected $F(5,43) = 1.74$, $p = .14$). Response accuracy was also analyzed in an ANOVA with fraction of artifact-free trials (i.e., 1st100, middle 100, and last 100) as a within-subjects factor. The main effect of number of trials was not significant (Huynh-Feldt corrected $F(2,19) = 2.84$, $p = .09$). This analysis suggests that, despite its effects on other aspects of behavior, time on task did not have a substantial influence on response accuracy.

EEG Analyses

Average spectra revealed changes in frontal theta and parietal alpha bands over time (Fig. 3). The changes in frontal midline theta (i.e., average power densities and peak amplitudes at Fz) were analyzed in a doubly multivariate ANOVA with time of measurement (i.e., 15 min periods) as a within-subjects factor. The main effect of time of measurement was significant (multivariate $F(18,178) = 2.05$, $p<.01$, $\eta^2 = .17$). Average power in the theta band increased from a mean of 199.36 $\mu V^2/Hz$ (SD = 97.50) in the first 15 min period to a mean of 256.58$\mu V^2/Hz$ (SD = 135.57) in the last 15 min period. Peak amplitude in the theta band increased from a mean of 272.4$\mu V^2/Hz$ (SD = 146.0) in the first 15 min period to a mean of 390.8$\mu V^2/Hz$ (SD = 227.1) in the last 15 min period. This analysis suggests that theta increased with time on task. Moreover, this analysis suggests that time on task had a progressive effect on frontal midline theta activity. Within-subjects contrasts showed significant linear increases in average theta power densities ($F(1,10) = 7.42$, $p<.01$, $\eta^2 = .48$) and in peak theta amplitudes ($F(1,10) = 9.31$, $p<.01$, $\eta^2 = .48$).

The changes in midline parietal alpha activity (i.e., average power densities and peak amplitudes at Pz) were analyzed in a doubly multivariate ANOVA with time of measurement (i.e., 15-min periods) as a within-subjects factor. The main effect of time of measurement was significant (multivariate $F(18,178) = 2.20$, $p<.005$, $\eta^2 = .18$). Average alpha power densities increased from a mean of 307.4 $\mu V^2/Hz$ (SD = 434.3) in the first 15 min period to a mean of 459.0$\mu V^2/Hz$ (SD = 593.9) in the last 15 min period. This analysis suggests that alpha increased progressively with time on task. Within-subjects contrasts showed significant linear increases in average alpha power densities ($F(1,10)6.07$, $p<.03$, $\eta^2 = .38$). Peak alpha amplitudes increased and trended similarly, but not significantly so ($F(1,10) = 4.11$, $p = .07$).

ERP Analyses

The N100, P200, and P300 latencies and amplitudes were analyzed in separate doubly multivariate ANOVAs with number of artifact-free trials (i.e., 1st100, middle 100, and last 100) as a within-subjects factor. For N100 and P300 alike, the main effect of number of trials was not significant (N100 multivariate $F(4,54) = 1.56$, $p = .20$; P300 multivariate $F(4,34) = 0.79$, p =

.54). This analysis suggests that time on task did not influence the N100 or P300. In the P300 range above 500 ms, P300 amplitudes in the last 100 trials were slightly larger than the 1st 100, but not significantly.

For P200 the main effect of number of trials was significant (multivariate $F(4,54) = 6.28$, $p<.0005$, $\eta^2 = .32$). This analysis suggests that time on task influenced the P200s. Moreover, this analysis suggests that time on task had both a linear and a curvilinear effect on the P200s. Within-subjects contrasts showed significant linear ($F(1,14) = 4.76$, $p<.05$, $\eta^2 = .25$) and quadratic ($F(1,14) = 18.4$, $p<.001$, $\eta^2 = .57$) trends, for the P200 amplitudes. Within-subjects contrasts did not show significant linear or quadratic trends for the P200 latencies ($p>.05$). Normalized P200 amplitudes averaged 0.55 (SD = .10) in the 1st 100 trials, 0.68 (SD = .10) in the middle 100 trials, and 0.45 (SD = .11) in the last 100 trials.

Classification

We applied our classification procedure to EEG recordings from 14 subjects (two subjects had too few EEG epochs for model estimation). The EEG epochs were synchronized with the onset of each math problem, extending from –5 s to +8 s relative to each stimulus onset. As such there was some overlap among the EEG segments. However a second analysis of 3.5 s segments with no overlap produced highly similar results, so we will focus only on the long-epoch results. We also reduced the likelihood of EMG artifact by low-pass filtering the EEG with 11- or 18-Hz cutoffs.

For each subject we constructed a KPLS model using either linear or Gaussian (nonlinear) kernels and selected the best model as described above. We then constructed a discretized linear regression classifier for each model, which served to classify the KPLS component scores for each EEG epoch. Results for linear and Gaussian kernels were not superior, and on average linear kernels had slightly better results, so we focus on linear kernels here. Classification accuracies (Fig. 4) across both classes for 18-Hz filtered EEG ranged from 91.12 to 100% (mean = 98.30, Table 2). The corresponding range for 11-Hz filtered EEG was 89.53 to 98.89% (mean = 98.30%). The number of KPLS components ranged from 1 to 4 (mean 2.77) for 18-Hz EEG and from 1 to 5 (mean 3.76) for 11-Hz EEG (Table 2). With as few as two components, the separation of classes was usually evident from the distribution of KPLS scores for single EEG epochs. The test-set data for the first- and last 15 min blocks occupied distinct regions in the space of the KPLS scores (Fig. 5).

The scalp topography of the KPLS weights can serve as an indicator of which regions or electrodes strongly influence classification. For example, by plotting the product of the KPLS model weights and the average EEG spectra in limited frequency bands in one subject (Fig. 6), we found that a broad set of fronto-central midline sites was important for classification in the theta band. In the alpha band, the discriminating electrodes were tightly concentrated over midline parietal site Pz.

KPLS Model Prediction

We also examined the predictive validity of the KPLS-DLR models by testing them with data from the first nine intervening 15 minute periods (between first and last). The behavior of the classifiers for these periods was consistent with an orderly, progressive migration of single-trial KPLS predictions from the non-fatigued to the fatigued class. This observation agrees with the trends we observed in response times, EEG measures, and behavioral observations. We examined these patterns of migration by inspecting graphs of the predicted scores for the first two components of the KPLS models for single subjects (Fig 7). Initially, the predicted points overla pped with the region occupied by the non-fatigued training set. Over time, the predicted points shifted towards the fatigue region. For each of 11 subjects, we computed the center of mass (mean) of the KPLS classification scores for each of 10 fifteen minute blocks (Figure 8). These means correspond to the centers of the clouds of points shown moving from the alert region to the fatigue region in Figure 7. Negative scores represent alert states and positive scores represent fatigue states. For most subjects there is a progressive shift of the means toward the fatigue state, with some intermittent reversals. Three subjects entered fatigue as early as block 1 (15-30 min) whereas one subject did not enter fatigue until block 7 (105-120 min). The majority of subjects (n=6) entered fatigue by block 3 (45-60 min).

DISCUSSION

Our results showed that the relative power of theta, alpha, and other EEG spectral measures co-varied with the degree of fatigue that subjects experienced. Specifically, theta and alpha activity increased across time on task. In general, the ERPs were not responsive to mental fatigue. With the exception of an unexpected increase in P200 amplitudes, none of the expected effects on other components were significant. We also found that we could accurately model and predict fatigue as a function of EEG measures using a statistical learning theory based classifier.

Behavioral Measures

Time on task produced decreased general activation (i.e., self-reported energy) and preparatory arousal (i.e., self-reported calmness) and increased general deactivation (i.e., self-reported tiredness) but did not influence moods. These effects support the assertion that our task produced a state of mental fatigue. Observed activity progressively increased while observed alertness progressively decreased over time. Moreover, there was a progressive, but moderate, slowing effect on response times. However, time on task did not influence response accuracy. Together, these results suggest that our subjects experienced mental fatigue, but did not sacrifice accuracy as may be expected if motivation had waned. The moderate, general increases in RT over time also indicate increasing mental fatigue, but not a severe increase as may be expected if lapses or sleep episodes had occurred frequently. Fatigue due to prolonged vigilance or drowsiness appears to have different effects on error rates than cognitive fatigue. Makeig and Inlow (1993) found that the EEG-based alertness estimate follows closely the time course of actual error rate. However, we found that cognitively fatigued subjects continued to perform with low and stable error rates for up to three hours. The main behavioral impact of cognitive fatigue appears to be a slowing of mental processes, as response times trend significantly and progressively higher over time.

EEG and ERP Measures

The EEG analyses suggest that time on task had a progressive influence on frontal midline theta and parietal alpha activity. Both rhythms increased as a function of time on task. Our inspection of the EEG spectra did not indicate effects outside the theta and alpha bands. In particular, there were no indications of effects at 14 Hz or in the beta band. Our results do not support an overall slowing of the EEG in mental fatigue, as much as they indicate specific increases in frontal midline theta and midline parietal alpha power. A detailed analysis of our classification results can provide more specific details for individual subjects, which we will report in the future.

ERP Measures

The ERP analyses suggest that time on task did not have substantial effects on N100, P200, and P300 amplitudes or latencies. The one exception was P200 amplitude, for which time on task had a curvilinear influence, with larger amplitudes in the middle of the task. There was no specific hypothesis about P200 and its sensitivity to fatigue in other contexts is poorly documented. The non-significant effect on P300 amplitudes was in the direction predicted by the "increasing workload" hypothesis. P300 amplitudes were larger during the periods of relatively high mental fatigue as compared to fresh performance.

For ERPs, a limiting factor in the success of statistical algorithms for modeling performance or cognitive states is the accurate estimation of wave shape in the presence of high noise or background EEG. Methods such as discrete wavelet decompositions and neural network pattern recognition allow for automated learning of models between the physiology and behavior, such as the features of ERPs that diagnose the transitions between normal and cognitive overload conditions. Kernel methods, such as KPLS allow for linear and nonlinear mappings of ERPs to cognitive states or performance states with minimal assumptions about the underlying mechanisms, and maximum adaptivity to individual differences. Although we employed such methods in this study, we did not find clear relationships between ERPs and fatigue states. Either the effects are too small to be detected or nonexistent in the present paradigm.

Classification and Prediction

KPLS-DLR classification of single trial EEG epochs was about 90% to 100% for with a mean of 97% to 98% depending on the low-pass cutoff. A small increase in classification accuracy appears to derive from including EEG in the 11-18 Hz range. The performance of these classifiers is highly accurate for single trials, and may serve as the basis for predictive models of mental fatigue in operational settings. Inspections of the predictive behavior of the KPLS models showed an orderly relationship between the scores and time on task as well as between the scores and correlated behavioral, subjective, and performance measures. In the 11 subjects that we examined for model predictions, our classifier generalized to new data from intervening blocks in an orderly manner, showing a progression from alertness to fatigue. Subjects fatigued at widely different rates, but most subjects entered the classifier region corresponding to fatigue by 45-60 min on task. We are continuing our assessment of the classifier with these data and are now testing classifier models that represent multiple states of alertness.

Makeig and Inlow (1993) performed a coherence analysis of fluctuations in moving-average measures of performance and EEG power to evaluate the nature and strength of the relationship between alertness and the EEG spectrum. Results were verified using linear regression, correlation, and error-sorted spectral averaging methods. These analyses provided evidence that changes in performance on an auditory detection task were linearly related to specific patterns of changes in the EEG power spectrum at all EEG recording sites and over a wide range of time scales. This relationship was variable between, but stable within subjects. The present study also shows that the individual EEG spectral signatures and patterns of change are useful for estimation of cognitive fatigue. The trend across subjects in EEG spectral change is explained by an alpha-theta model, with both alpha and theta being greater in fatigued states than in normal states. However, the estimation of momentary cognitive fatigue states from single EEG trials of 13 s duration was most effective with multi-electrode frequency spectra of individual subjects. The KPLS algorithm correctly identified features of EEG spectra for each individual, and accurately classified over 90% of single EEG trials as having come from a period of low or high cognitive fatigue in most subjects.

Possible concerns about time-locked estimates of EEG power spectra

In our analyses of EEG effects and our classification of EEG epochs, the EEG power spectral estimates that we used were from epochs that were synchronized with the task-relevant stimulus (appearance of the equation). This raises two possible concerns: a) that our spectra reflect the effects of time-locked activity such as ERPs or movement-related potentials, b) that the EEG spectra we measured represent activity close in time to the presentation and performance of the mental arithmetic problems. We think that ERP or movement-related potentials did not bias our EEG spectra for four reasons: a) the ERP effects were insignificant or inconclusive, b) the 13-s epochs are very long as compared to ERPs which usually resolve in 1 s or less, c) a separate analysis of readiness potentials associated with responses (not described) did not yield significant time-on-task effects, d) a separate classification analysis (not described) showed that neither time locked ERP time series nor readiness potential time series were useful for classification of mental fatigue. Nevertheless it is possible that a focused analysis of ongoing EEG epochs, which are not synchronized with task events, would lead to different conclusions. We speculate that such analyses may lead to reduced EEG effects and less accurate classification from increased temporal separation between the performance of mental arithmetic and the EEG spectral estimates.

Future work will examine details of the KPLS models to describe individual frequency/electrode effects. For operational applications, we will also develop methods for minimizing the number of electrodes in the models, testing predictions of the models with new experiments, and developing adaptive statistical classifiers for on-line use.

References

Barth, J.L., Holding, D.H., & Stamford, B.A. (1976). Risk versus effort in the assessment of motor fatigue. *Journal of Motor Behavior, 8,* 189-194.

Belyavin, A., & Wright, N.A. (1987). Changes in electrical activity of the brain with vigilance. *Electroencephalography and clinical Neurophysiology, 66,* 137-144.

Brown, I.D., Tickner, A.H., & Simmonds, D.C.V. (1970). Effect of prolonged driving on overtaking criteria. *Ergonomics, 13,* 239-242.

Dinges, D.F. (1995). An overview of sleepiness and accidents. *Journal of Sleep Research, 4 (Suppl. 2),* 4-14.

Duta, M., Alford, C., Wilson, S., & Tarassenko, L. (2004). Neural network analysis of the mastoid EEG for the assessment of vigilance. *International Journal of Human-Computer Interaction, 17(2),* 171-195.

Fairclough, S.H., Venables, L., & Tattersall, A. (2004). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology, 56,* 171-184.

Fatigue Resource Directory (2000). Remarks by Jim Danaher. In National Transportation Safety Board Proceedings: Opening Remarks. Retrieved October 4, 2005, from http://human-factors.arc.nasa.gov/zteam/PDF_pubs/NTSB_Proceedings/NTSB_proceedings_opremarks.pdf

Florian, G., & Pfurtscheller, G. (1995). Dynamic spectral analysis of even-related EEG data. *Electroencephalography and clinical Neurophysiology, 95,* 393-396.

Galbraith, G.C., & Wong, E.H. (1993). Moment analysis of EEG amplitude histograms and spectral analysis: Relative classification of several behavioral tasks. *Perceptual and Motor Skills, 76,* 859-866.

Gevins, A., Leong, H., Du, R., Smith, M.E., Le, J., DuRousseau, D., et al. (1995). Towards measurement of brain function in operational environments. *Biological Psychology, 40,* 169-186.

Gevins, A., & Smith, M.E. (1999). Detecting transient cognitive impairment with EEG pattern recognition methods. *Aviation, Space and Environmental Medicine, 70,* 1018-1024.

Gevins, A., & Smith, M.E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex, 10,* 829:839.

Gevins, A., Smith, M.E., Leong, H., McEvoy, L., Whitfield, S., Du, R., et al. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors, 40,* 79-91.

Gevins, A., Smith, M.E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex, 7,* 374-385.

Gevins, A.S., Zeitlin, G.M., Ancoli, S. & Yeager, C.L. (1977). Computer rejection of EEG artifact. II. Contamination by drowsiness. *Electroencephalography and Clinical Neurophysiology, 42*, 31-42.

Gundel, A., & Wilson, G.F. (1992). Topographical changes in the ongoing EEG related to the difficulty of mental tasks. *Brain Topography, 5,* 17-25.

Haga, S., Shinoda, H., & Kokubun, M. (2002). Effects of task difficulty and time-on-task on mental workload. *Japanese Psychological Research, 44,* 134-143.

Hankins, T.C., & Wilson, G.F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environmental Medicine, 69,* 360-367.

Horne, J.A. (1988). Sleep loss and "divergent" thinking ability. *Sleep, 11(6),* 528-536.

Humphrey, D.G., & Kramer, A.F. (1994). Toward a psychophysiological assessment of dynamic changes in mental workload. *Human Factors, 36*, 3-26.

Isreal, J.B., Chesney, G.L., Wickens, C.D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. Psychophysiology, 17, 259-273.

Isreal, J.B., Wickens, C.D., Chesney, G.L., & Donchin, E. (1980). The event-related brain potential as an index of display-monitoring workload. *Human Factors, 22(2),* 211-224.

Jung, T., Makeig, S., Stensmo, M., & Sejnowski, T.J. (1997). Estimating alertness from the EEG power spectrum. *IEEE Transactions on Biomedical Engineering, 44,* 60-69.

Kennedy, J.L. (1953). Some practical problems of the alertness indicator. In W. F. Floyd & A. T. Welford (Eds.), Symposium on Fatigue (pp. 149-153). Oxford, England: H.K.Lewis & Co.

Kramer, A.F., Sirevaag, E.J., & Hughes, P.R. (1988). Effects of foveal task load on visual-spatial attention: Event-related brain potentials and performance. Psychophysiology, 25, 512-531.

Kramer, A.F., Trejo, L.J., & Humphrey, D. (1995). Assessment of mental workload with task-irrelevant auditory probes. *Biological Psychology, 40,* 83-100.

Kramer, A.F., Trejo, L.J., & Humphrey, D.G. (1996). Psychophysiological measures of workload: Potential applications to adaptively automated systems. In R. Parasuraman & J. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 137-162). Mahwah, NJ: Lawrence Erlbaum Associates.

Kramer, A.F., Wickens, C.D., & Donchin, E. (1983). An analysis of the processing requirements of a comples perceptual-motor task. Human Factors, 25, 597-621.

Laukka, S.J., Järvilehto, T., Alexandrov, Y.I., & Lindqvist, J. (1995). Frontal midline theta related to learning in a simulated driving task. *Biological Psychology, 40,* 313-320.

Lorist, M.M., Boksem, M.A.S., & Ridderinkhof, K.R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cognitive Brain Research, 24,* 199-205.

Lorist, M.M., Klein, M., Nieuwenhuis, S., De Jong, R., Mulder, G., & Meijman, T.F. (2000). Mental fatigue and task control: Planning and preparation. *Psychophysiology, 37,* 614-625.

Makeig, S., & Inlow, M. (1993). Lapses in alertness: coherence of fluctuations in performance and EEG spectrum. *Electroencephalography and clinical Neurophysiology, 86,* 23-35.

Makeig.S., & Jung, T. (1995). Changes in alertness are a principal component of variance in the EEG spectrum. *NeuroReport, 7,* 213-216.

Makeig, S., Jung, T., & Sejnowski, T.J. (1995). Using feedforward neural networks to monitor alertness from changes in EEG correlation and coherence. *Advances in Neural Information Processing Systems, 8,* 931-937.

Mangun, G.R., & Hillyard, S.A. (1990). Allocation of visual attention to spatial locations: Tradeoff functions for event-related brain potentials and detection performance. Perception & Psychophysics, 47, 532-550.

Mast, T.M., & Heimstra, N.W. (1964). Effects of fatigue on vigilance performance. *Journal of Engineering Psychology, 3(3),* 73-79.

McCarthy, G., & Wood, C.C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology, 62,* 203-208.

Mecklinger, A., Kramer, A.F., & Strayer, D.L. (1992). Event-related potentials and EEG components in a semantic memory search task. Psychophysiology, 29, 104-119.

Montgomery, L.D., Montgomery, R.W., & Guisado, R. (1995). Rheoencephalographic and electroencephalographic measures of cognitive workload: analytical procedures. *Biological Psychology, 40,* 143-159.

Neri, D.F., Shappell, S.A., & DeJohn C.A. (1992). Simulated sustained flight operations and performance, Part 1: Effects of fatigue. *Military Psychology, 4(3),* 137-155.

Nikolaev, A.R., Ivanitskii, G.A., & Ivanitskii, A.M. (1998). Reproducible EEG alpha-patterns in psychological task solving. *Human Physiology, 24,* 261-268.

Nyenhuis, D.L., Stern, R.A., Yamamoto, C., Luchetta, T., & Arruda, J.E. (1997). Standardization and validation of the Visual Analogue Mood Scales, Clinical Neuropsychologist, 11, 407-415.

Pellouchoud, E., Smith, M.E., McEvoy, L., & Gevins, A. (1999). Mental effort-related EEG modulation during video-game play: Comparison between juvenile subjects with epilepsy and normal control subjects. *Epilepsia, 40,* 38-43.

Rosipal, R., Trejo, L.J., Matthews, B. (2003). Kernel PLS-SVC for linear and nonlinear classification. In T. Fawcett and Nina Mishra (Eds.), *Proceedings of the Twentieth International Conference on Machine Learning*, (pp. 640-647), Washington DC, Aug. 21-24.

Russo, M.B., Stetz, M.C., & Thomas, M.L. (2005). Monitoring and predicting cognitive state and performance via physiological correlates of neuronal signals. *Aviation, Space, and Environmental Medicine, 76,* C59-C63.

Shingledecker, C.A., & Holding, D.H. (1974). Risk and effort measures of fatigue. *Journal of Motor Behavior, 6,* 17-25.

Šípoš, I. (1971). Verbal vigilance performance and visual fatigue. *Studia Psychologica, 13,* 141-142.

Sirevaag, E.J., Kramer, A.F., Coles, M.G., & Donchin, E. (1989). Resource reciprocity: an event-related potentials analysis. Acta Psychologica, 70, 77-97.

Smit, A.S., Eling, P.A.T.M., & Coenen, A.M.L. (2004). Mental effort causes vigilance decrease due to resource depletion. *Acta Psychologica, 115, 35-42*.

Smith, M.E., Gevins, A., Brown, H., Karnik, A., & Du, R. (2001). Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction. *Human Factors, 43,* 366-380.

Smith, M.E., McEvoy, L.K., & Gevins, A. (2002). The impact of moderate sleep loss on neurophysiologic signals during working-memory task performance. *Sleep, 25,* 784-794.

Stedman's Medical Dictionary. (2005). Accessed online on October 4, 2005 at http://www.stedmans.com/.

Stern, R.A. (1997). Visual Analogue Mood Scales. Odessa, FL: Psychological Assessment Resources.

Thayer, R. E. (1986). Activation-Deactivation Adjective Check List: Current overview and structural analysis. *Psychological Reports, 58*, 607-614.

Thayer, R. E. (1989). *The Biopsychology of Mood and Arousal*. New York: Oxford University Press.

Trejo, L.J., Kramer, A.F., & Arnold, J.A. (1995). Event-related potentials as indices of display-monitoring performance. *Biological Psychology, 40,* 33-71.

Trejo, L.J., & Mullane, M.M. (1995). ERPs and visual signal detection performance: Classification functions based on wavelet decompositions. *Psychophysiology, 32,* S77.

Trejo, L.J, & Shensa, M.J. (1999). Feature extraction of event-related potentials using wavelets: An application to human performance monitoring. *Brain and Language, 66,* 89-107.

Ullsperger, Metz, & Gille, (1988). The P300 component of the event-related potential and mental effort. Ergonomics, 31, 1127-1137.

van der Hulst, M., Meijman, T., & Rothengatter, T. (2001). Maintaining task set under fatigue: a study of time-on-task effects in simulated driving. *Transportation Research Part F: Traffic Psychology & Behavior, 4,* 103-118.

van der Linden, D., Frese, M., & Sonnentag, S. (2003). The impact of mental fatigue on exploration in a complex computer task: Rigidity and loss of systematic strategies. *Human Factors, 45,* 483-494.

van der Linden, D., Frese, M., & Meijman, T.F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica, 113,* 45-65.

Venturini, R., & Lytton, W.W., & Sejnowski, T.J. (1992). Neural network analysis of event related potentials and electroencephalogram predicts vigilance. In J. Moody, S. Hanson, & R. Lippmann (Eds.). *Advances in Neural Information Processing Systems 4* (pp. 651-658). San Mateo, CA: Morgan Kaufmann.

Vidulich, M.A., Stratton, M., Crabtree, M., & Wilson, G. (1994). Performance-based and physiological measures of situational awareness. *Aviation, Space, and Environmental Medicine, 65,* A7-A12.

Wickens, C., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information processing resources. Science, 221, 1080-1082.

Wilson, G.F., & Fisher, F. (1991). The use of cardiac and eye blink measures to determine flight segment in F4 crews. *Aviation, Space and Environmental Medicine, 62,* 959-962.

Wilson, G.F., & Fisher, F. (1995). Cognitive task classification based upon topographic EEG data. *Biological Psychology, 40,* 239-250.

Wilson, G.F., Fullenkamp, P., & Davis, I. (1994). Evoked potential, cardiac, blink, and respiration measures of pilot workload in air-to-ground missions. *Aviation, Space, and Environmental Medicine, 65,* 100-105.

Wilson, G.F., & Russell, C.A. (2003a). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Human Factors, 45 (3),* 381-389.

Wilson, G.F., & Russell, C.A. (2003b). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors, 45 (4)*, 635-643.

Author Note

We would like to acknowledge the support of the NASA/ ASEE Faculty Fellowship and the NASA Ames Educational Associates programs. We also acknowledge the contributions of Richard W. Montgomery and Bernadette Luna for theoretical discussions and use of the mental arithmetic task, and Melina L. Trejo for analyzing and scoring video tapes of subject’s performances.

36

Table 1. Scale used to quantify observed activity and alertness.

Value	Activity	Alertness
1	Little or no movement (0-10% time)	Asleep
2	Occasional movement (10-25% time)	Dozing off but still responding
3	Intermittent movement (25-50% time)	Distracted, yawning or only somewhat alert
4	Frequent movement (50-75% time)	Completely awake and mostly alert
5	Constant movement (75-100% time)	Completely awake and fully alert

37

Table 2. KPLS-DLR classification accuracies by filter cutoff and class membership.

Low pass cutoff	TPC Train	TPC Test	TPC Class 1	TPC Class 2	Mean number of components
11 Hz	99.96	97.01	97.37	95.57	3.76
18 Hz	99.86	98.30	98.78	96.97	2.77

Figure Captions

Figure 1. Schematic diagram of events in the mental arithmetic task.

Figure 2. Effects of time on task on response time and accuracy. Response times trended linearly upwards over time beginning after Block 3, or 45 minutes on task. Error rates declined over time but the trend was not significant.

Figure 3. Average EEG spectra across all subjects for the first (fine line) and final (heavy line) 15 min blocks of the math task. Left: Electrode Fz shows an increase in theta power near 6-7 Hz. Right: electrode Pz shows an increase in alpha power near 8-11 Hz.

Figure 4. Classification accuracies for 14 subjects and the averages across subjects with 18 Hz band pass. Light grey bars show test set accuracies for overall classification; dark grey bars are for Class 1 (alert) epochs, and white bars are for Class 2 (fatigued) epochs.

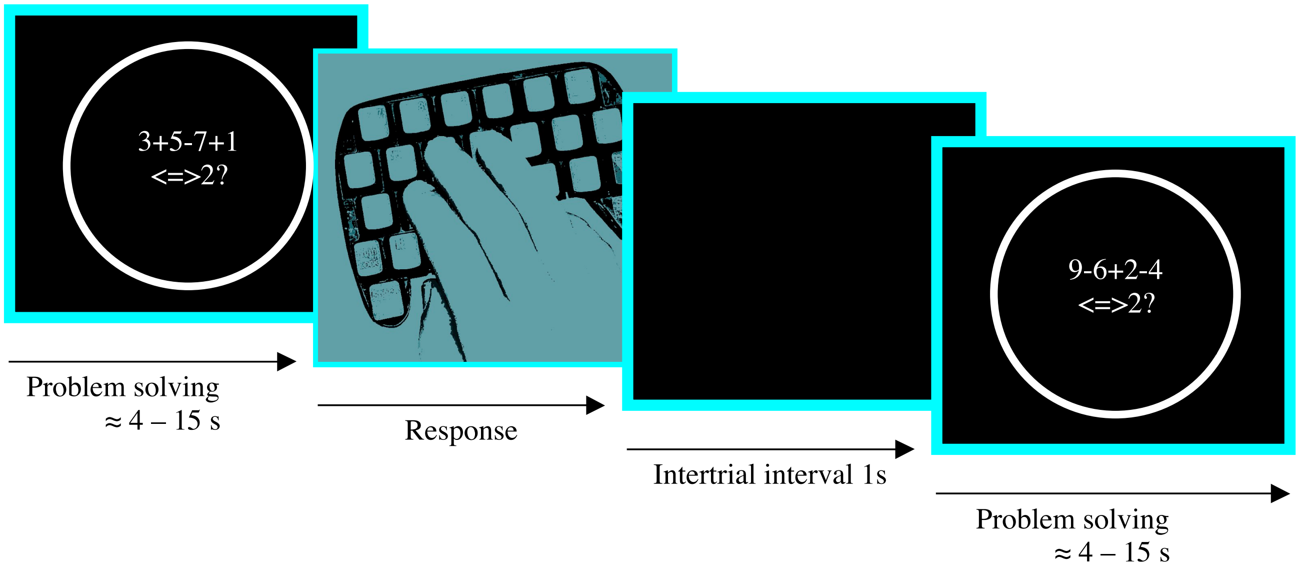
Figure 5. Example of KPLS scores predicted for single-trial EEG spectra for early (green/dark circles) and late (yellow/light triangles) blocks of the mental arithmetic task in one subject.

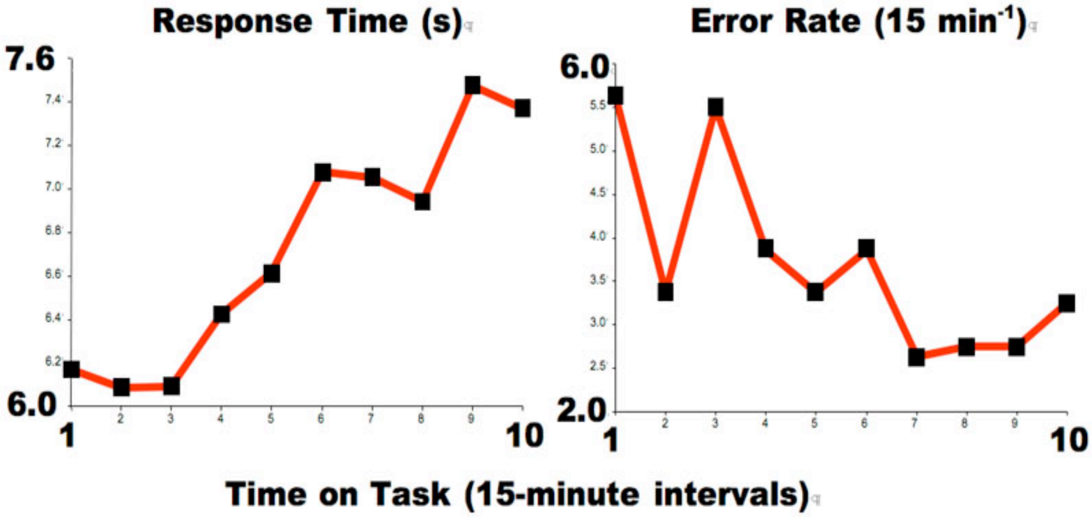
Figure 6. Topographical maps fit to the first KPLS component weights in theta and alpha bands for one subject (819). The colored areas are smoothed normalized absolute values, with the largest values in red and the smallest in blue.

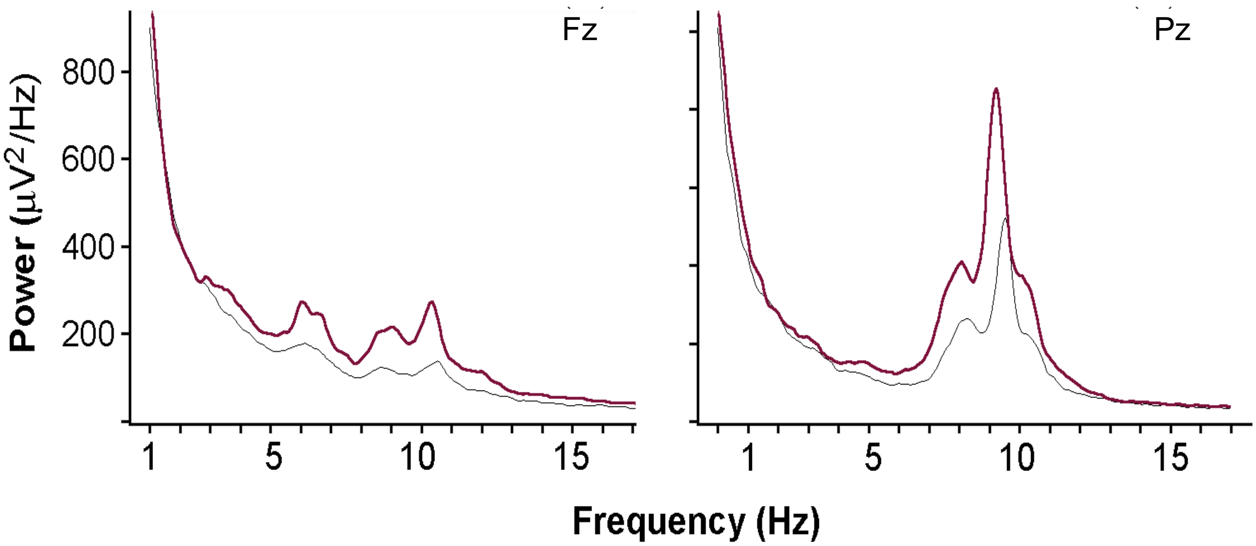
Figure 7. Estimating the development of fatigue over time in one subject (819). KPLS scores were predicted for EEG epochs from nine 15 minute blocks between the training set blocks (B1 & B12). Block 2 = 15-30 min, block 3 = 30-45 min, block 4 = 45-60 min, … , block 10 = 135-150 min. Black circles and purple crosses are the KPLS C1 and C2 scores of single EEG epochs from fatigued (block 1) and non-fatigued (block 12) training sets, respectively. Colored diamonds are the KPLS C1 and C2 scores (x = C1, y = C2) of single EEG epochs for intervening 15-minute blocks 2-10. In this subject, the drift of the orange diamonds in block 3 away from the

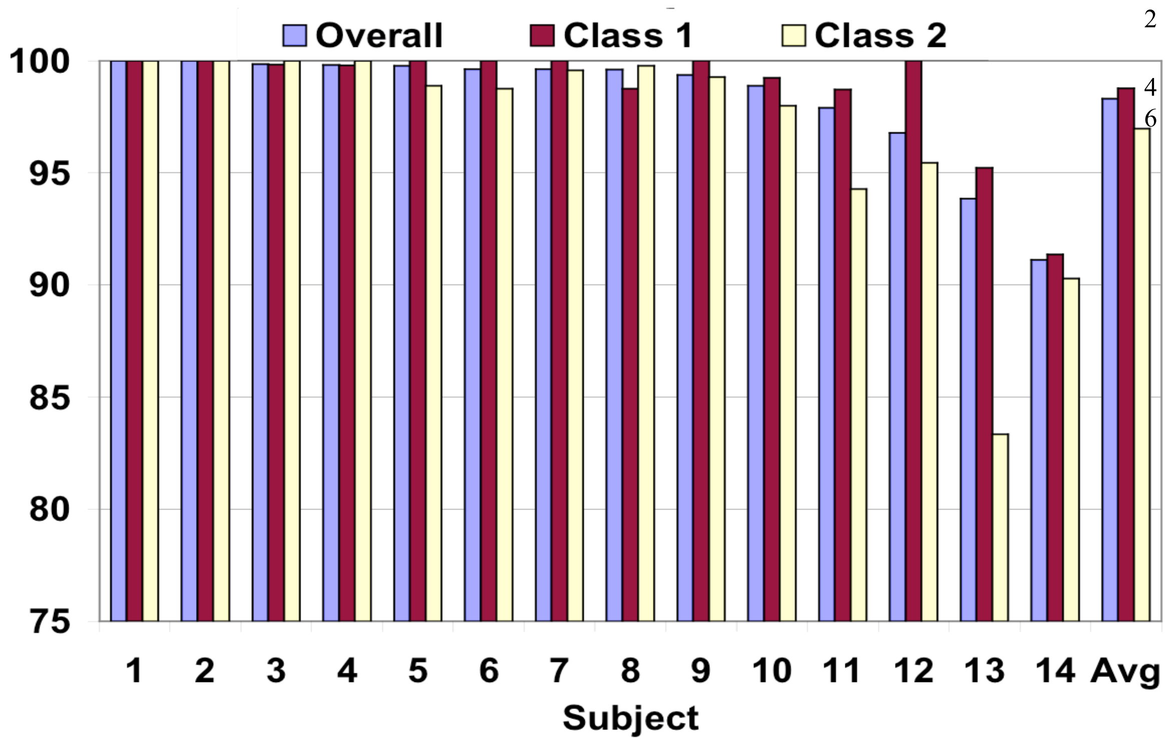
black circles and towards the purple crosses marks the onset of fatigue after 30-45 minutes on task. By the tenth block most predicted scores fell in the fatigue region, as defined by the training data set.

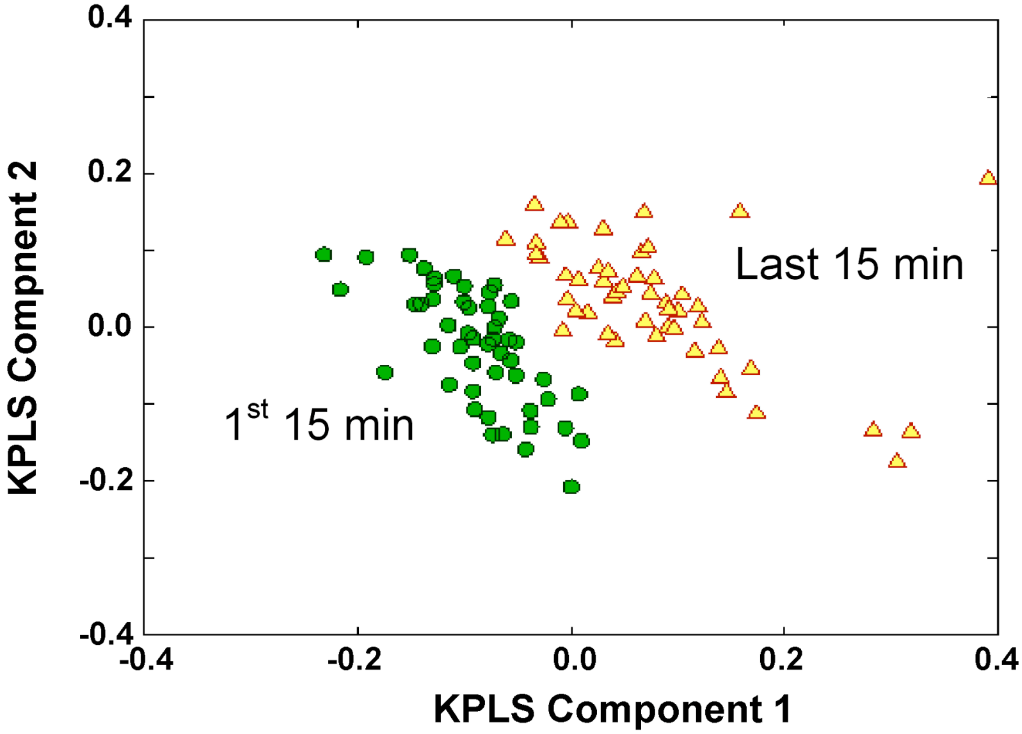
Figure 8. Means of the KPLS classification scores for each of 10 fifteen minute blocks in 11 subjects. These means correspond to the centers of the clouds of points shown moving from the alert region to the fatigue region in Figure 7. Negative scores represent alert states and positive scores represent fatigue states.


3+5-7+1
<=>2?
9-6+2-4
<=>2?
Problem solving
≈ 4 – 15 s
Response
Intertrial interval 1s
Problem solving
≈ 4 – 15 s

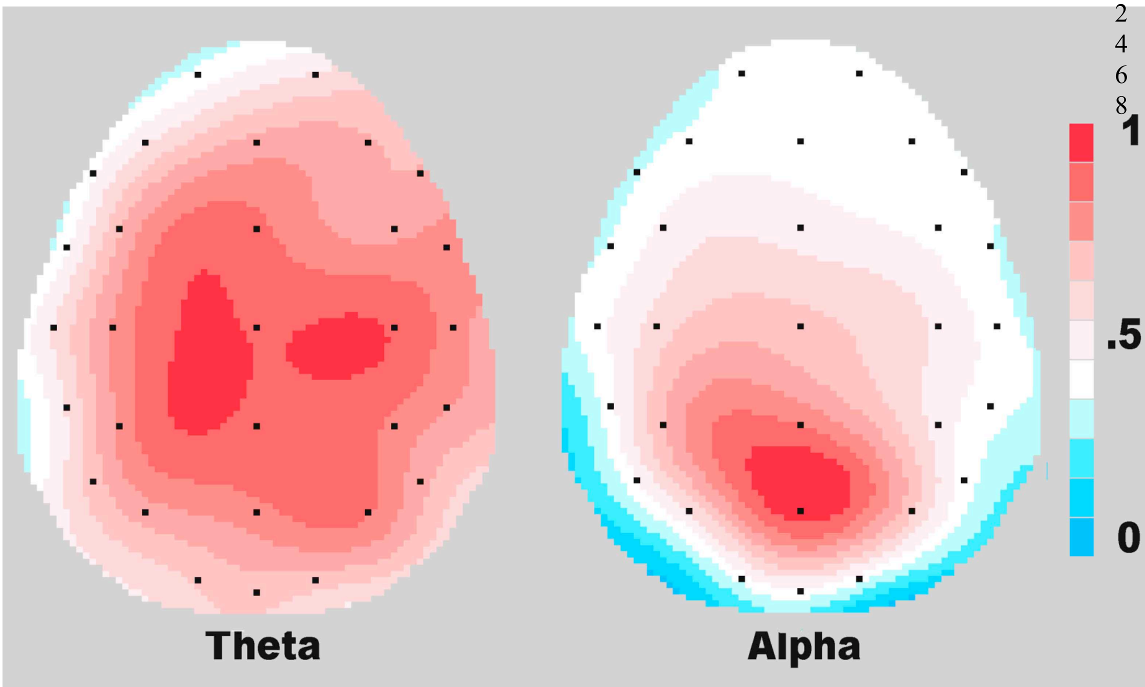

Response Time (s)
7.6
7.4
7.2
7.0
6.8
6.6
6.4
6.2
6.0
1
2
3
4
5
6
7
8
9
10
Error Rate (15 min-1)
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1
2
3
4
5
6
7
8
9
10
Time on Task (15-minute intervals)

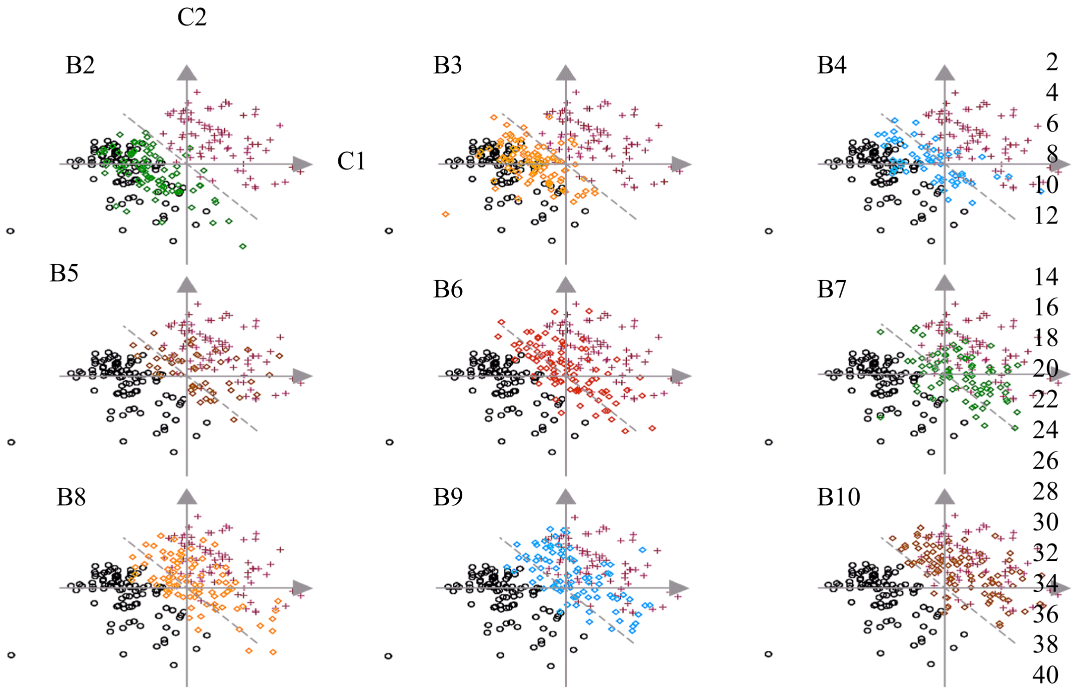

Fz
Pz
Power (μV2/Hz)
800
600
400
200
1
5
10
15
Frequency (Hz)

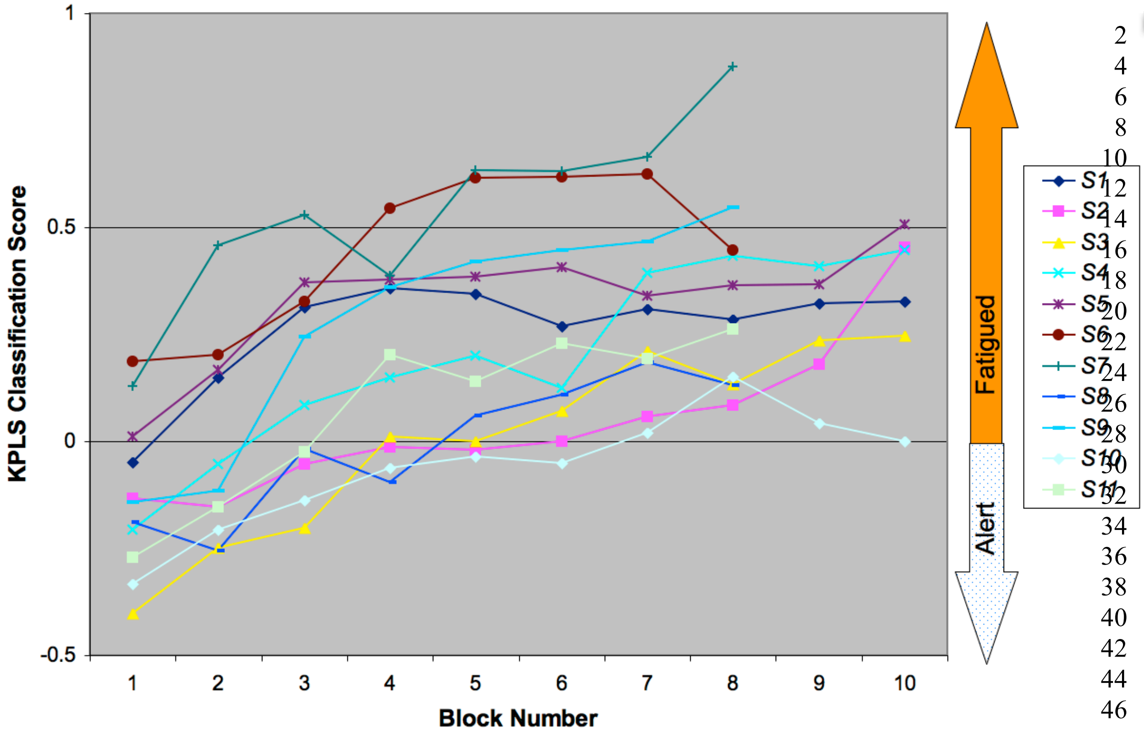

Overall
Class 1
Class 2
100
95
90
85
80
75
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Avg
Subject


0.4
0.2
0.0
-0.2
-0.4
KPLS Compnent 2
-0.4
-0.2
0.0
0.2
0.4
KPLS Component 1
Last 15 min
1st 15 min

1


2
4
6
8
1
.5
0
Theta
Alpha


C2
B2
C1
B3
B4
B5
B6
B7
B8
B9
B10


KPLS Classification Score
1
0.5
0
-0.5
1
2
3
4
5
6
7
8
9
10
Block Number
Fatigued
Alert
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11